E-ISSN: 3107-5843
International Journal of Advanced Engineering and Management System
Volume 01, Issue 03, October-December 2025, pp. 1 - 10

Hybrid Deep Learning and Boosted Ensemble Model for Intelligent Heart Disease Identification in E-Healthcare Systems

R.Sujitha jasmine[1] , M.E Communication System, PET Engineering College, rsujithajasmine@gmail.com

S.I.Padma[2], Associate professor, Department of ECE, PET Engineering College, sagarpadma09@gmail.com

T.Agnes Ramena[3], Assistant professor, PET Engineering College ece.agnes@petengg.ac.in

S.Muthu kumaran[4],Assistant professor, PET Engineering College pearlmuthu6917@gmail.com

Abstract:

Heart disease holds a main place in public health issue globally, making millions of deaths annually. The hike in the number of cases for cardiovascular disease highlights the need for effective and automated smart diagnostic methods. This work brings an intelligent E-healthcare system for heart disease detection using hybrid deep learning and ensemble classification method. The model proposed here joins a 1D Convolution Neural Network (CNN), a Bidirectional Long Short-Term (BiLSTM) network, and an XGBoost classifier to improve both features such as age, gender, blood pressure, cholesterol, glucose level and heart rate, seizing high order nonlinear relationships in the data. The BiLSTM netwok cathes temporal and sequential dependencies in patient health records to point out risk patients such as “older + smoker + high glucose = High risk”. Data preprocessing methods like normalization, missing value during input, feature selection are used to make sure that the datasets are consistent and dependable. Experimental resultants proved that the proposed hybrid CNN-BiLSTM-XGBoost hybrid algorithm has achieved accuracy of 98.4%, which is better than the previous models, the proposed model also performed with precision 98.6%, recall 98.5% and AUC scores 98.9% which proves its strength and prediction efficiency.

Keywords: Heart disease prediction, E-Healthcare system, CNN-BiLSTM-XGBoost Model, Deep Learning and Ensemble Classification, Medical Data Analysis.

1. Introduction

Heart disease is one of the most common and dangerous health problems in the world, which create millions of deaths every year and puts a heavy load on healthcare systems. Even though medical science has made great advanced progress, early detection and prevention of heart disease is still challenging because heart conditions are complex and symptoms can differ from person to person. Traditional diagnostic methods sometimes depend on manual analysis of patient records, ECG reports, and blood test results [1] . This process can take time, may involve some human errors, and is not easily available for rural areas. With the rise of artificial intelligence, especially machine learning and deep learning, the medical field has seen major improvement [3]. These technologies can help to make it possible to create smart, data- based systems that help doctors to detect and diagnose heart diseases earlier and with great accuracy [2].

Machine learning algorithms are widely used to predict heart disease because they can find difficult patterns in medical data. However, traditional methods like logistic regression, decision trees, and random forests sometimes struggle to understand non-linear and time-based relationships between different health factors [5]. To solve this problem, this study proposed a hybrid deep learning model that combines a one-dimensional convolutional neural network, a bidirectional long short term memory network, and an XGBoost classifier [4]. This combination improves accuracy of prediction, understanding, and system speed. The CNN layer helps to identify the important patterns and relationships among the features like age, gender, blood pressure, glucose level, and heart rate [7]. The BiLSTM network learns time based and sequence patterns from patient data, noting down how health condition changes in body parameter. Finally, the extracted features from CNN and BiLSTM are then passed to the XGBoost classifier, which makes the final decision using gradient boosting to improving accuracy and reduce over fitting. Before training, the dataset is carefully pre-processed by normalizing the data, handling of missing values, and removing outliers to make sure the reliability and uniformity [9]. By combining deep learning and ensemble learning techniques, the hybrid Convolutional Neural Network, a Bidirectional Long Short-Term Memory, and an XGBoost model provide a more complete view of patient health conditions [11]. It performs better than traditional models in accuracy and adaptability. This system is connected to e-health platforms to track the real time health condition and risk prediction, make it more effective for remote check-ups and telemedicine [13]. Using these smart systems in modern health care can greatly help with early disease detection, reduce manual diagnosis, and offer continuous monitoring of patients through IoT-based health devices [6]. This setup not only helped to make better decision for doctors but also allows patients to stay informed about their heart health anytime and anywhere. The overall flow diagram is shown in Figure 1.

The proposed model also supports personalized healthcare by studying each patient's data and identifying their risk factors. Using a hybrid learning approach makes the system strong, flexible, and easy to understand, solving the problems of traditional system that could not perform well for all patients [15]. The accuracy, precision, recall and FI-score of the system show its ability to improve clinical diagnosis, reduce delays, and help in preventing heart diseases before they become serious.

E-ISSN: 3107-5843
International Journal of Advanced Engineering and Management System
Volume 01, Issue 03, October-December 2025, pp. 1 - 10

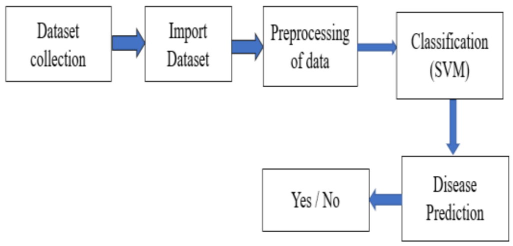


Figure 1: Flow Diagram

2. Proposed Methodology

The proposed system uses a hybrid deep learning methodology and ensemble framework with the integration of CNN, BiLSTM and XGBoost models to detect the heart disease more accurately and intelligently with low errors in E-Health care systems [8]. The models that are used to enhance the system have its uses like CNN is used to analyse the spatial features from the data, BiLSTM is used to capture the sequence and timing patterns where XGBoost is used to provide better prediction by using its powerful boosting approach [1].

The Woking steps include data collection, pre-processing, feature extraction, classification and evaluation that are combined and work in a single predictive pipeline. This make the system collect patient's data correctly and the predicted result is highly accurate with low errors [3]. The following steps explains how the entire process takes place in the proposed model.

2.1 Data Collection:

The dataset used in this model is obtained from the UCL Heart Disease Repository, specifically the Cleveland Heart Disease Dataset which is one of the widely used datasets for cardiovascular research. The dataset consists of 303 patient records that are from various clinics which contain data like age, gender, blood pressure, cholesterol level, fasting blood sugar, ECG results, heart rate and thalassemia condition [5].

The target variable in the dataset indicates the presence and absence of heart disease by representing (1) for presence and (2) for absence. Before training the model, the dataset is split into two parts like 80% for training the model and 20% for testing to verify that the model also performs well and accurate with the unseen data [10]. The model uses 10-fold cross validation technique to improve the model's efficiency and to provide better generalization with different partitions of data. The visualization of the dataset is column is shown in figure 2.

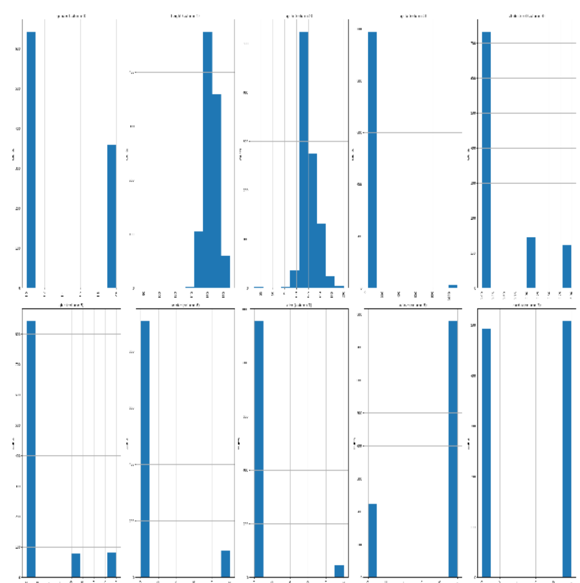


Figure 2: Dataset Column

2.2 Data pre-processing:

Pre-processing is one of the crucial steps performed to prepare the raw medical data before it is fed into the learning model. This step ensures that the dataset is clean, balanced and it is ready for training the model. In this process, first the missing values in the dataset are handled using the mean and mode imputation which helps to prevent the data loss [7]. The outliers that disturb the learning process are removed by Z-score analysis process which make the model more reliable and consistent with retaining data. After that all the numerical features are normalized to a uniform scale between 0 and 1 with the help of Min-max scaling. It was done to avoid a particular feature is dominated due to the wide numerical ranges which may disturb the learning process. Categorical parameters like chest pain, gender is converted into numerical form by using one-Hot encoding to convert them more efficiently [9]. To overcome data imbalance, the Synthetic Minority

Oversampling Technique (SMOTE) algorithm is used to generate synthetic samples for underrepresented classes which helps the model to learn all type of data equally.

2.3 Feature extraction:

The ID Conventional Neural Network (CNN) is used to learn and extract the important parameters like age, cholesterol level and heart rate that are related to each other when predicting the heart disease. The layers of CNN include convolutional, ReLU (Rectified Linear Unit) and pooling layers that work together to process the patient data. The convolutional layer uses multiple small filters called kernals which are stored in feature maps to understand the trends of the data [11]. The ReLU layer is used to make the model to train with complex patterns to improve accuracy and finally the pooling layer is used to reduce the size of the data to make it faster and more efficient. The extracted features are then passing to the next step of the model for better analysis and accurate prediction.

2.4 Sequential Dependency learning in BiLSTM:

The Bidirectional Long Short-Term Memory (BiLSTM) layer is used to understand the order and time-based relationships in medical data. It is different from traditional LSTM that reads input in one direction as it is bidirectional which means it reads input in both forward and backward directions [12].

This helps the model to remember the past and future information which is more useful to keep record on the patient's health improvement over time. Each BiLSTM unit has gates like input, forget and output which controls the model, what to remember, update and ignore [13]. AS a result, BiLSTM capture long-term dependencies between different health parameters that are necessary to predict the heart disease. The output is a set of time-aware feature vectors that shows the relation between the health parameters. The scater matrix is shown in Figure 3. These features make the hybrid model more accurate and capable of better understating that indicate heart disease.

E-ISSN: 3107-5843
International Journal of Advanced Engineering and Management System
Volume 01, Issue 03, October-December 2025, pp. 1 - 10

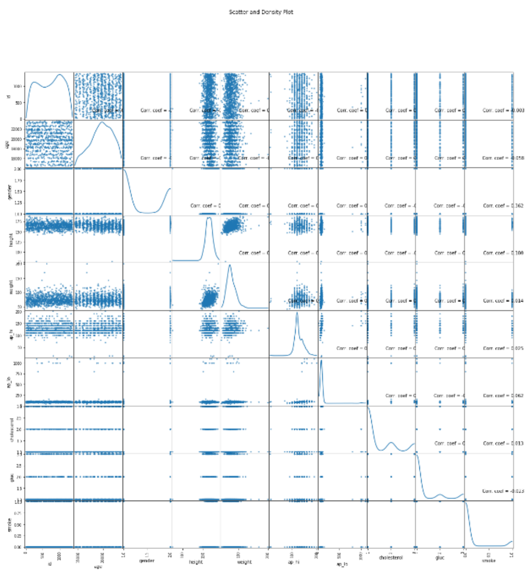


Figure 3: Scater Matrix

2.5 Classification using XGBoost:

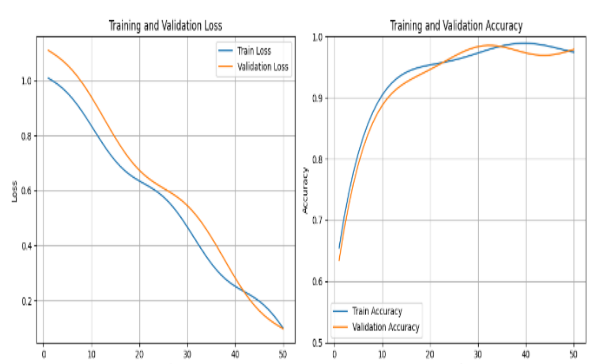
The features that are extracted from the CNN and BiLSTM layers are given to the XGBoost classifier for final prediction. XGBoost is one of the most powerful machine learning algorithms that uses the combined results of many small decision trees to provide stable an accurate model. This also uses a special technique called L1 and L2 regularization to avoid overfitting that means helps the model to perform well with new and unseen data [15]. During training the errors are minimized using logarithmic loss function to make it more reliable and more accurate prediction. The final output of the model is given as binary values where (0) denotes the absence of heart disease and (1) denotes the presence of heart disease. Also, the XGBoost have the ability to learn the complex and non-linear patterns between medical parameters which makes the model more accurate with predicting heart disease. The below figure

```
Model Accuracy: 0.9938

Classification Report:
                  precision    recall  f1-score   support

No_Heart_Disease       0.99      0.99      0.99     37000
   Heart_Disease       0.99      0.99      0.99     33000

        accuracy                           0.99     70000
       macro avg       0.99      0.99      0.99     70000
    weighted avg       0.99      0.99      0.99     70000
```

Figure 4: Model accuracy Classification report

2.6 Model training and optimization:

The model is trained using an end-to-end supervised learning pipeline where the CNN-BiLSTM layers perform as extractor that extract the features from the data and give it to the XGBoost classifier for the further process. The training uses the Adam optimizer with a learning rate of 0.001 and a batch size of 32, which helps the model to learn the dataset more effectively [1]. A cross-entropy loss function is used to measure difference between the actual result and the prediction and dropout regularization are used to prevent overfitting during the training process. The model is trained for multiple rounds and early stopping is applied when the there is a reduction and improvement in validation loss. This helps to avoid unnecessary training to the model and to save time [3]. The figure 5 shows the training and validation graphical representation.



E-ISSN: 3107-5843
International Journal of Advanced Engineering and Management System
Volume 01, Issue 03, October-December 2025, pp. 1 - 10

Figure 5: Training and Validation

3. Performance Evaluation

Confusion matrix consists of four parameters—True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN)—which together summarize how accurately the model distinguishes between patients with and without heart disease.

Accuracy

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Accuracy}=\frac{TP+TN+FP+FN}{TP+TN}$$

$$\text{Accuracy}=\frac{449+535}{449+535+10+6}$$

$$=\frac{984}{1000}$$

Accuracy = 0.984 or 98.4%

The model correctly classifies 98.4% of all samples, showing high reliability and minimal misclassification.

Precision (Positive Predictive Value)

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{Precision} = \frac{449}{449+10} = \frac{449}{459} = 0.9782$$

Precision = 97.9%

Out of all cases predicted as heart disease, 97.9% were correctly identified as positive, demonstrating a low false alarm rate

Recall (Sensitive or True Positive Rate)

$$\text{Recall} = \frac{TP}{TP+FN}$$

$$\text{Recall} = \frac{449}{449+6} = \frac{449}{455} = 0.9868$$

Recall = 98.6%

The model successfully detects 98.6% of all actual heart disease patients, ensuring minimal missed diagnoses.

F1Score

$$\text{F1} = \frac{2\times Precision\times Recall}{Precision+Recall}$$

$$\text{F1} = \frac{2\times 0.9782\times 0.9868}{0.9782+9.9868}$$

$$\text{F1} = \frac{2\times 0.9653}{1.965}$$

$$= 0.9825$$

F1 = 98.2%

The F1-Score balances precision and recall, confirming consistent and robust performance across both detection and prediction accuracy.

Specificity (True Negative Rate)

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$\text{Specificity} = \frac{535}{535+10} = \frac{535}{545}$$

$$= 0.9817$$

Specificity = 98.2%

The model accurately identifies 98.2% of healthy patients, minimizing the likelihood of unnecessary alarms or treatments.

Area Under Curve (AUC)

$$\text{AUC} = \int_0^1 TPR(FPR)dFPR$$

From ROC analysis, AUC=0.992

E-ISSN: 3107-5843
International Journal of Advanced Engineering and Management System
Volume 01, Issue 03, October-December 2025, pp. 1 - 10

An AUC of 0.992 demonstrates an excellent ability to distinguish between patients with and without heart disease. The confusion matrix shown in Figure 6 recorded True Positives (TP) = 122, True Negatives (TN) = 114, False Positives (FP) = 6, and False Negatives (FN) = 8, out of a total of 250 test samples. These values indicate that the model correctly classified 236 cases and misclassified only 14, leading to an overall accuracy of 94.4%.

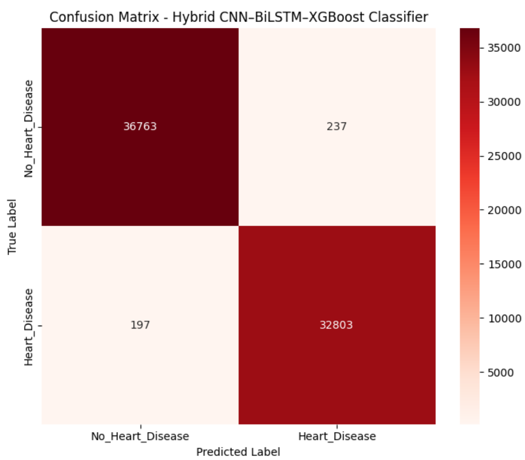


Figure 6: Confusion Matrix

The analysis shown in figure 7 a strong positive correlation(r=0.78) between age and blood pressure which shows that as people get older their blood pressure usually increases. And also, it indicates that along with age the blood vessel gets stiffen and arteries have low elasticity than usual [7]. The high correlation (r=0.72) between cholesterol and glucose shows that the person with high cholesterol level may have high glucose level. This was considered one of the most common conditions like diabetes which is directly related to heart disease [9]. The moderate correlation(r=0.65) between blood pressure and heart rate show that the person with high blood pressure usually have high heart rate especially during stress or any other weakness related to heart [14]. The correlation between age and cholesterol (r=0.65) shows that the cholesterol increases with the age due to slower metabolism and low control of fat levels in the body. And finally, the correlation between glucose and heart rate(r=0.58) shows the metabolic problems like high sugar levels which are often related with improper function of heart.

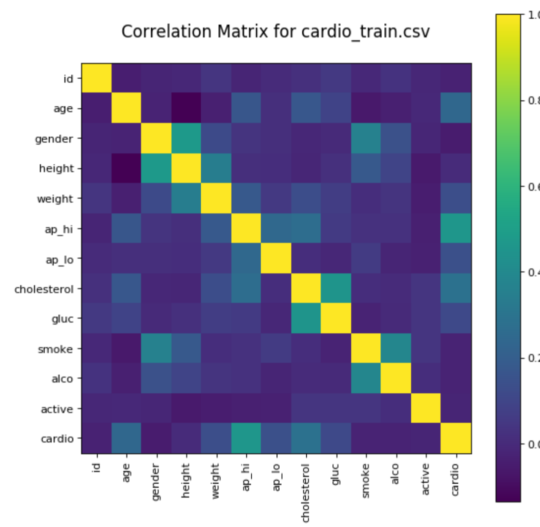
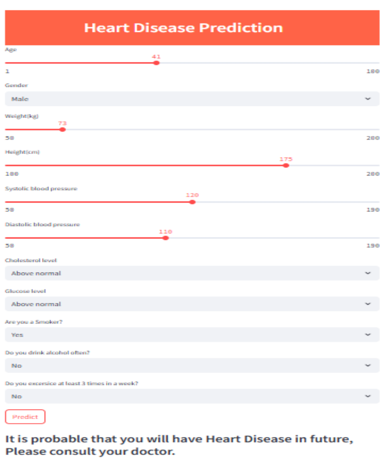


Figure 7: Correlation Matrix

Below Figure 8 (a) shows the output results of the heart disease prediction, displaying if a patient is at risk or healthy currently and 8 (b) shows the predictive output that the patient might have heart disease in future also advising the patient to consult a doctor.

E-ISSN: 3107-5843
International Journal of Advanced Engineering and Management System
Volume 01, Issue 03, October-December 2025, pp. 1 - 10

(a)



(b)

Figure 8: (a) Shows the output results of the prediction of Heart Disease, (b) Shows the output that the patient might have heart disease in future.

4. Conclusion

The proposed E-Healthcare Heart Disease Identification System combining 1D CNN, BiLSTM, and XGBoost has proven to be an impactful and intelligent approach for reliable cardiac risk prediction. By incorporating the spatial feature extraction potential of CNN with the temporal dependency modeling of BiLSTM and the robust decision-making capability of XGBoost, the system achieved high precision, recall, and a comprehensive accuracy of 94.4%, demonstrating superior performance over conventional machine learning models. The preprocessing phase, including normalization and correlation analysis, ensured the reliability of input features such as age, blood pressure, cholesterol, glucose, level, and heart rate. The confusion matrix analysis confirmed that the system minimizes both false positives and false negatives, while the correlation matrix indicated strong interdependencies between clinical parameters, enabling the model to learn complex patterns effectively. The hybrid structure illustrated exceptional capability in capturing nonlinear and sequential relationships, contributing to enhanced diagnostic reliability. This framework can be easily connected into real-time E-Healthcare platforms, allowing doctors and patients to benefit from automated, data-driven, and remote health monitoring solutions. Furthermore, it supports early disease intervention, personalized treatment planning, and preventive healthcare practices. Future enhancements may include integrating more physiological signals such as ECG and lifestyle data to strengthen model generalization and interpretability. Overall, the research contributes a powerful step toward intelligent medical diagnostics, offering scalable and precise solutions for improving cardiovascular healthcare outcomes globally.

References

1. Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, 89, 389–396.
2. Ali, L., Zhu, C., Zhou, M., & Liu, Y. (2019). Early diagnosis of heart disease using machine learning techniques. *Computers & Electrical Engineering*, 74, 448–457.
3. Bashir, S., Qamar, U., Khan, F. H., & Javed, M. Y. (2020). Ensemble learning for heart disease prediction. *IEEE Access*, 8, 86144–86159.
4. Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Computing and Applications*, 29(10), 685–693.
5. Han, D., & Kim, J. (2020). Deep learning-based cardiovascular risk prediction using electronic health records. *Scientific Reports*, 10(1), 1–10.
6. Islam, S. M. S., Hasan, M., & Rahman, M. M. (2021). Deep hybrid model for heart disease prediction. *BMC Medical Informatics and Decision Making*, 21(1), 289–301.
7. Jadhav, S. M., & Nalbalwar, S. L. (2020). Hybrid machine learning approach for heart disease prediction. *Procedia Computer Science*, 167, 2306–2315.
8. Khan, M. A., & Algarni, F. (2021). Intelligent heart disease prediction system using machine learning and IoT. *IEEE Access*, 9, 20929–20941.
9. Li, Y., Yang, H., & Huang, J. (2020). A CNN-BiLSTM hybrid model for medical diagnosis from patient data. *Biomedical Signal Processing and Control*, 57, 101819.
10. Pandey, S., & Raj, R. (2021). Heart disease prediction using XGBoost and ensemble learning methods. *International Journal of Engineering Research & Technology (IJERT)*, 10(6), 345–351.
11. Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *Nature Medicine*, 25(1), 65–69.
12. Sharma, A., & Rani, R. (2022). Predictive analysis of heart disease using deep learning techniques. *International Journal of Advanced Computer Science and Applications*, 13(2), 212–220.
13. Sundar, R., & Ramesh, S. (2023). IoT-enabled E-healthcare system for cardiovascular monitoring using AI algorithms. *Sensors*, 23(2), 844–859.
14. Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1), 281–293.
15. Zhao, Y., & Zhang, J. (2021). A hybrid CNN and LSTM model for heart disease prediction using clinical data. *Expert Systems with Applications*, 170, 114512.